DOCUMENT RESUME

ED 143 852 08 CE 012 479

TITLE	The Design of Career Education Evaluations in 1975-76.
INSTITUTION	New Educational Directions, Crawfordsville, Ind.
SPONS AGENCY	Office of Career Education (DHEW/OE), Washington, D.C.
PUB DATE	Jul 77
GRANT	G007604329
NOTE	43p.; For related documents see CE 012 414-415 and CE 012 479
EDRS PRICE	MF-$0.83 HC-$2.06 Plus Postage.
DESCRIPTORS	Adult Education; *Career Education; Data Analysis; Data Collection; Elementary Secondary Education; *Evaluation Methods; Federal Aid; Formative Evaluation; *Program Evaluation; Program Validation; Summative Evaluation; *Technical Writing
IDENTIFIERS	Third Party Evaluation

ABSTRACT

Based on reviews of the final performance and evaluation reports of 81 career education projects funded by the U.S. Office of Education during 1976, this comprehensive report seeks to identify ways of making future evaluation efforts in this area more fruitful. The introduction consists of definitions of key terms and a clarification of the source of information on which this report is based. Part I is a discussion of common pitfalls in career education evaluation design, execution, and reporting. Part II presents several unique and/or high quality (according to the author(s)) evaluation strategies identified in reviewed reports. Part III, addressed specifically to project directors, concerns the effective utilization of contractual evaluation services. (Author/BM)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

NED NEW EDUCATIONAL DIRECTIONS, INC.
a nonprofit service organization
BOX 307 CRAWFORDSVILLE, INDIANA 47933 (317) 362-8877

ED143852

THE DESIGN OF CAREER EDUCATION EVALUATIONS IN 1975-76

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

July 1977

Prepared under Grant Number G007604329 from:

The U.S. Office of Career Education
Office of Education
Department of Health, Education, and Welfare

Project Title: Synthesizing and Communicating Career Education Evaluation Results

Project Director: Deborah G. Bonnet, Director
Research and Evaluation Programs

Other Publications in this Series:

Checklist for Reporting Results of Student Outcome Studies in Career Education

Recommendations for the Implementation and Management of Career Education Projects

What Does Career Education Do For Kids: A Synthesis of 1975-76 Evaluation Results

CE 012 479

DISCRIMINATION PROHIBITED

Title VI of the Civil Rights Act of 1964 states: "No person in the United States shall, on the ground of race, color, or national origin, be excluded from participation in, be denied the benefits of, or be subjected to discrimination under any program or activity receiving Federal financial assistance." Title IX of the Education Amendments of 1972, Public Law 92-318, states: "No person in the United States shall, on the basis of sex, be excluded from participation in, be denied the benefits of, or be subjected to discrimination under any education program or activity receiving Federal financial assistance." Therefore, career education projects supported under Sections 402 and 406 of the Education Amendments of 1974, like every program or activity receiving financial assistance from the U.S. Department of Health, Education, and Welfare, must be operated in compliance with these laws.

The material in this publication was prepared pursuant to a grant from the Office of Education, U.S. Department of Health, Education, and Welfare. However, points of view or opinions expressed do not necessarily represent policies or positions of the Office of Education.

TABLE OF CONTENTS

	page
Introduction	1
Projects included in the review	1
Evaluation	2
Three types of evaluation	3
Part I--Critique of Common Evaluation Practices	5
Overall Evaluation Designs	5
Reporting	9
Questionnaires and Interviews	11
Sampling for Student Impact Assessment	12
Research Designs	16
Descriptions of Evaluated Programs	19
Outcome Measurement Instruments and Strategies	20
Statistical Analyses and Interpretation	23
Part II--Unique Solutions to Common Evaluation Problems	28
Referenced Projects	33
Part III--A Note to Project Directors on How to Get the Most From Your Third Party	35

INTRODUCTION

Career education evaluation has received a great deal of attention in the past several years. Its importance has been emphasized clearly and its methodologies have been discussed widely. This attention has paid off; career education, though still a relatively new concept, now is probably among the most extensively evaluated of educational movements.

However, room remains for improving career education program evaluation practices. Since evaluation is likely to retain its high priority for several years to come, it seems appropriate at this time to examine recent career education evaluations in order to identify ways of making future efforts in this area more fruitful. This was one purpose of our review of final performance and evaluation reports from 81 K-12 career education programs.

The remainder of this section consists of definitions of key terms and a clarification of the source of information on which this report is based. Part I is a discussion of common pitfalls in career education evaluation design, execution, and reporting. Part II presents several unique and/or high-quality evaluation strategies identified in reviewed reports. Part III addressed specifically to project directors, concerns the effective utilization of contractual evaluation services.

Projects Included in the Review

The project of which this report is one product was designed to encompass 108 career education programs funded by the U.S. Office of Education during the 1975-76 academic year. The projects were funded under two separate federal programs; 47 received three-year grants from the Bureau of Occupational and Adult Education pursuant to Section 104(c), Part D of PL 90-567 and 61 were funded for one year by the Office of Career Education (OCE) under Section 406, Title IV of PL 93-380. All Part D projects were designed to incorporate grades K-12, K-14, or K-adult. The OCE projects included those in funding Categories 1 (K-12 incremental improvement), 2 (senior high school settings), 3 (special populations) and one program for adults funded in Category 5.

However, all 108 projects were not included in the review: Some had not yet submitted their final performance and evaluation reports as of our cut-off date of March 4, 1977, even though the large majority of the reports were due September 30, 1976. A few reports which were reviewed were not included in the analyses discussed in this report because the programs were not designed to develop and test strategies for career educating students; one such program was for the development of a state plan for career education which was funded under OCE's Category 1. Another report in this series, *What Does Career Education Do For Kids: A Synthesis of 1975-76 Evaluation Results* addresses only the projects reporting student impact evaluation results at the K-12 level which met the criteria for inclusion in that synthesis.

The number of projects falling under each of these circumstances is shown below:

	OCE	Part D	Total
Proposed for review	61	47	108
Final performance and evaluation reports received by NED as of 3/4/77.	49	32	81
Addressed in this report	45	32	77
Met criteria for inclusion in synthesis of student impact evaluation results.	21	26	47

Evaluation

The term *evaluation* is defined very broadly here to include activities usually considered *measurement* or *assessment*.

We have made no distinctions between evaluative data and descriptive data partially because what may be purely descriptive to one person could be evaluative to another. For example, a report of the number of school staff members trained through a project may in some circumstances be an indication of the project's success in sparking local interest in career education whereas in other cases the same data would have no meaningful evaluative implications.

Another reason for using a broad definition of evaluation is that data which alone are only descriptive often serve as crucial components of evaluation systems and lead to evaluative conclusions when interpreted in conjunction with other components of the systems.

Program evaluation has many purposes and serves the needs of a number of audiences, but among its most important functions is its role in the responsibility of exemplary projects to develop model programs suitable for replication in other settings. This primary mission of all exemplary or demonstration projects entails not only the development of programs and lines of communication with potential adopters of the programs, but also includes providing enough information about the model program to allow others to make informed decisions of whether it should be adopted in their schools. This means that the potential adopter needs and deserves answers to the question, "Does this program work?", but the program's effectiveness is of little concern if the question, "What is the program?" cannot be answered. Thus, documentation of a program's effects is not sufficient; the probable causes of the effects also must be documented if exemplary programs are to have value for schools other than those where they were developed. Since precise and quantitative descriptions of the tasks and resources involved in installing a model program and in implementing it are important components of the evaluation system, it is important to consider these components in analyzing career education evaluation.

Three Types of Evaluation

Evaluation is generally considered to be of two types--process and outcome or formative and summative. However, we have found three categories to be more meaningful for externally-funded career education projects. The typical sequence of events for such projects is: 1) The funded project staff members develop and implement various strategies designed to bring about 2) educational reform. These changes in instructional practices and in the relationships between the school and community, in turn, have 3) an affect on students. Depending on how it is viewed, (2) above could be considered either a process or an outcome; thus, the use of three levels of evaluation rather than the usual two. Evaluation strategies are categorized here as 1) *project strategy assessment,* 2) *educational reform outcome assessment,* or 3) *student outcome assessment* on the basis of the level at which evaluation data have the most direct bearing. Thus, student achievement results are classified as student outcome assessment data even though they may reflect the effectiveness of

project strategies and the extent of educational reform. The illustration below demonstrates a typical sequence of events and the level associated with each.

Level	Event
Project Strategies	Project staff develop training program. Teachers are trained.
Educational Reform Outcome	Teachers develop positive attitudes toward career education. Teachers develop career education knowledge and skills. Teachers implement career education activities with students (Students experience career education activities).
Student Outcome	Students develop new skills, attitudes, and knowledge.

Categorizing events and their associated evaluation strategies in this manner looses meaning in cases where the funded staff members assume a major role in the direct delivery of career education experiences to students, such as in most experience-based career education (EBCE) programs. However, in the vast majority of projects encompassed by this review the primary roles of the funded staff were to design programs and to influence other educators and community members to implement them; some contact with students generally was maintained but this was rarely central to the staff's responsibilities.

PART 1

CRITIQUE OF COMMON EVALUATION PRACTICES

Overall Evaluation Designs

Evaluation designs were in most cases multi-faceted, addressing numerous process and outcome objectives. Virtually every project reported at least one evaluation activity concerned with project strategies and educational reform outcomes and the great majority (88% of Part D and 78% of OCE) included student impact assessment data. Furthermore, student impact assessments tended to be comprehensive, focusing on several objectives at each of several grade levels. A typical student impact evaluation consisted of assessments of three or four objectives at each of three or four grade levels. Nevertheless, the strategies and resources leading to success or failure in achieving student outcome objectives tended to be inadequately defined.

Inadequate information about project strategies and activities. A major concern of potential adopters of exemplary programs and of other audiences of final reports is the logistics of project management, including such matters as the skills required of project staff members and the equipment and facilities needed for project operation. Information about project strategies and activities is interesting in itself and also serves as a background for the interpretation of outcome evaluation results.

Table 1 lists fourteen aspects of management which are pertinent to all of the 77 projects included in our review and which are likely to influence a project's success in effecting educational reform and student impact. As demonstrated in the table, many of these project management variables rarely are mentioned in final reports; even fewer reports than indicated in the table address these topics thoroughly.

We must point out, however, that the standards NED implicitly set for final reports by identifying the kinds of information we expected to find in them are *our* standards. The projects generating the reports had no access to them. The federal government had not endorsed them, nor have they yet, nor are they likely to. Although our first reaction to the finding that less than half of the reports indicated the number of positions on the project staff was one of alarm, this initial response was tempered by the realization that nothing in the instructions indicated that this information may be of interest to the readers of final reports.

Both Part D and OCE instructions for preparing performance reports are by necessity in the form of outlines which identify broad topics defined in general terms. It would behoove project staffs and evaluators alike to recognize that these instructions are the result of a complex series of compromises among several government agencies and that they represent only minimum requirements.

An indication of the typical report's failure to quantify major project activities or to address the quality of those efforts is seen in discussions of school staff development in career education, which was a major function of 76 of the 77 projects.

Only 33 of the project performance reports and 9 evaluation reports discussed either the content or the logistics of staff training. The total number of training sessions offered by the project was indicated in the project performance and/or evaluation report in 21 cases, but an unduplicated count of the school staff members involved in training was given for only 10 projects and in only 3 cases was the average amount of training per staff member indicated. Slightly over half of the projects' evaluations included assessments of the quality of at least one training session either in terms of its affect on participants' knowledge and attitudes or in terms of their satisfaction with the session. However, total staff development programs, which usually consist of multiple training sessions employing several training modes, were assessed qualitatively in only 15 cases. Data indicating the quantity and quality of other project functions such as community education and materials development are similarly sparse.

Inadequate information about career education activities resulting from the project's efforts. The weakest element of evaluation designs was usually in the identification of the amount and type of career education activities taking place. This problem took two forms.

First, projects' educational reform outcomes rarely were assessed thoroughly; only five reports provided clear indications of which groups contributed to the program's development and implementation, in which ways, and to what degree. This is particularly ironic for OCE programs in Category 1, as their main focus was to be on demonstrating incremental quality improvement defined in terms of educational reform outcomes.

Second, student impact evaluations tended to be unclear concerning the amount and type of career education experienced by the students whose outcomes were measured. Thus, the effectiveness of many student activities and programs was reported without those activities and programs being defined.

TABLE 1

PROJECT MANAGEMENT

Percentages of project and evaluation reports which discussed *any* aspect of the following components of project management:

	Project	Evaluator
Listing of funded staff positions	39%	16%
Qualifications or backgrounds of funded staff	17%	8%
Funded staff training	25%	1%
Coordination and communication among funded staff	26%	4%
Location of project's office(s) (e.g., in a school)	19%	5%
Project facilities (e.g., office space, equipment)	3%	3%
Fiscal management policies	3%	1%
Cost analysis beyond the required financial status report	9%	14%
Chronological plan for program development	57%	12%
Methods of announcing the availability of project services	78%	18%
Strategies for securing participation in project activities	25%	8%
Logistics and/or topics of school staff training	43%	12%
Composition and/or accomplishments of advisory committee	55%	14%
Obstacles or problems encountered	45%	23%

The result of this weakness is that evaluation designs were often fragmented, consisting of discrete sets of data which had little bearing on one another. For example, an evaluation may consist of a description of a career education curriculum guide, participants' assessments of a staff development program in the guide's proper use, and measurement of changes in students' decision-making skills. If evaluation results indicate that trainees were satisfied with the quality of the staff development program but that students' decision-making skills did not improve, it could be inferred that the activities suggested in the curriculum guide are ineffective. But this could be entirely incorrect, because we would not know whether the staff development program led to the actual use of the curriculum guide, whether it was used as intended, or whether it was used in the instruction of the students who were tested. Thus, we would not know whether to revise the curriculum guide, the training program, or both; all we know is that the staff were satisfied with the training program and that students' decision-making skills did not improve.

Lack of transportable recommendations. As evidenced in *Recommendations for the Implementation and Management of Career Education Projects,* project staffs and evaluators offered a number of transportable recommendations, as well as ones directed specifically to the project's locale. Nevertheless, many evaluations did not include this final step in the evaluation process. Some resulted in specific recommendations for the evaluated project, as they should, but did not address the needs of potential adopters of the exemplary program. Several evaluators made comments such as, "Since the project was not refunded for next year, no recommendations are offered for the program's improvement", which seem to indicate that the evaluation was intended to serve only local needs and to serve them only so long as career education was supported through special funds.

Reporting

Insufficient information. By far the most glaring weakness of career education evaluation is the quality of reporting and the worst problems are ones of omission. Evaluation reports, on the whole, fail to provide enough information about either the program under evaluation or about the evaluation itself to allow the reader to interpret and use the results. The problem is evident in the findings presented above and will also be addressed later. The *Checklist for Reporting Results of Student Outcome Studies in Career Education,* another document in this series, provides guidelines for reporting on student impact evaluations and was designed to alleviate that part of the problem with evaluation reporting. Therefore, we will not dwell on topics addressed in the *Checklist* and will try to keep our comments on omissions relating to other kinds of evaluation brief.

Organization. Another common, but by no means universal, weakness of evaluation reports is their organization. All too popular is the professional journal format, where the report begins with a discussion of all of the evaluation questions, proceeds to descriptions of all of the instruments, then to the evaluation groups, data-collection procedures, results, and, finally, the conclusions. This format, although very reasonable for a four-page journal article concerned with one research problem, can be rather awkward in a thirty-page evaluation report addressing ten evaluation questions. We are not going to suggest a new standard format because the point we wish to make is that there is no one organization that will cover all cases optimally. However, each of the following bases for organization has worked well in some cases:

* Sections for project strategies, educational reform outcome, and student impact assessment.
* Each program objective or component, or each evaluation question or hypothesis addressed separately. (So long as there is not a huge number of them.)
* Each data source separately (Such as all teacher data together, all third-grade student data together, etc.).

Three simple rules of thumb of report organization are:

* Avoid the need for redundancy. There were cases where a single set of data was presented and discussed in two or more places because it related to more than one section of the report as it was organized.
* Place related results in close proximity. For example, discussions of the type and amount of career education experienced by sophomores at Central High School should be easily associated with their outcome results. Pre-test results should be in the same table or at least on the same page as post-test results. If a table or the text refers to questionnaire item 8, it should not be necessary to consult the appendices to find out what the question was.
* Include a table of contents.

References to non-appended appendices are remarkably common.

Tables which do not indicate what the numbers within them represent were also encountered.

Internal inconsistencies within evaluation reports take the forms of conflicting data (e.g., two different counts of the number of advisory council meetings); inconsistent titles (of people, tests, etc.); and conflicts between tabulated data and discussions of them (e.g., the comparison group scored higher than the career education group according to the table, but results are discussed as if they were positive.)

Poor articulation between the evaluation report and the project's performance report. The two reports were often redundant, particularly in cases where both reports discussed the activities associated with each objective. Still, inconsistencies of the types listed above were common between the two reports and important information was often present in neither report. Better coordination between evaluators and project staffs would make the two reports more complementary and the package more thorough. Some redundancy should be retained, perhaps in the form of each report containing an abstract of the other, but there is little advantage in the evaluator's "seconding" everything the project says, or vice-versa.

Statements that data were collected, but no results given. Sometimes the reason for omitting the results of a particular evaluation strategy was given. It may be that a thorough report was prepared previously and was not included in the final report for the sake of brevity. In these cases a synopsis should still be included if for no other reason than to assure the reader that the motive for the omission is pure. The same should be done if the data are not presented because they were

judged to be of interest only to the project staff.

Negative evaluation findings should not be covered up. Besides the ethical issue, negative findings are useful and should be conveyed to others so that they may avoid the same mistake. Negative findings can also enhance the credibility of the positive ones.

Questionnaires and Interviews

Almost every evaluation involved the use of tailor-made instruments for gathering data such as community members' attitudes toward career education, educators' degree of implementation of career education programs, or students' opinions of career education activities. For the most part these evaluation components were well-conceived and well-executed, but a few errors were committed repeatedly.

Poorly worded questions or response options. For example, this question is ambiguous because it defines neither "career education activity" nor the time frame for the answer:

How many career education activities have you carried out?

This item, though perfectly clear to the adult-level reader, may have been over the heads of the junior high students who were asked it:

> Indicate your feelings about the time allowed for adequately completing laboratory activities.
> a. too much
> b. just right
> c. not enough

The problem with this one is not so much the question itself, but the meaning of the answer:

> Do you understand that career education encompasses *all* education: professional, technical, vocational?
> a. yes
> b. no

Incomplete information. Questionnaire results should be accompanied by a list of the questions and an indication of the group to whom it was administered, the number who completed it, the date of administration, and, if applicable, the return rate. The same types of information are needed with interviews. The interview guide should be included and the methodology should be described, at least to the extent of indicating whether interviews were conducted by telephone or in person.

Analyses of results. Most questionnaire and interview data should be reported in terms of the responses to each question; total "scores" are meaningful only in special cases such as with some attitudinal questionnaires. In deciding how much detail to report the object is to give enough information to make meaningful interpretations without bombarding the reader with page after page of numbers. One extreme to avoid is illustrated by a case where a 20-item opinion survey was given to 45 individuals. Results were averaged across respondents *and* items and reported simply as "81% positive". The other extreme is to report the responses to each question for each of ten or so groups separately. Information is lost by combining some of those groups, but the information that remains is more easily interpreted. If results are presented for each of several groups (such as elementary, middle school, and high school staffs), the number of people giving each answer should be converted to percentages to facilitate comparisons across groups.

Open-ended questions are good in some respects, but the results are difficult to deal with. If they are used, time and resources must be allowed for organizing the responses and identifying trends within them. The other alternatives, both of which were taken, are to present lengthy lists of verbatim remarks or to report simply, "The evaluation team interviewed thirty-two members of the community resource pool and learned that, all in all, they are enthusiastic about the program."

Sampling for Student Impact Assessment

A sample is a subset of a population. Samples are used in most research and evaluation activities because testing everyone in the population is expensive, inconvenient, and/or impossible. So that the results of research can be used to make predictions for the population, procedures are used to ensure that the sample is a subset of the population and that it resembles the population as much as possible. Therefore, a fundamental consideration in sample selection is, "To what population do we want to generalize?" In order to answer this question we must first establish the purpose of the evaluation and also of the program.

Exemplary or demonstration projects such as career education projects funded by the USOE have one primary mission: to develop programs which can be adopted or adapted by other schools to fulfill needs common to many or all American school systems. This mission implies many responsibilities for exemplary programs, one of them being to provide potential adopters of their program with sound bases for deciding whether to try them in their school systems. That is, a school administrator in California should be able to predict whether a program developed in Arkansas will work in his or her school on the basis of the Arkansas program's evaluation results.

This is equivalent to saying that the students who are chosen for the Arkansas evaluation should be representative of a national population, such as American high school students, gifted fifth-graders, or rural low-income kindergarteners. One way--in some respects the best way--to draw a representative national sample for testing the program's effectiveness is to choose randomly from all American rural low-income kindergarteners forty or so children to undergo the career education program. To the best of our knowledge, this has never been done. Fortunately, there are other ways to meet the assumption inherent in all research that the sample represents the population to which its outcomes are to be generalized. One of the most feasible ways for field research activities to meet this assumption involves a somewhat backward logic--students are "selected" for the exemplary career education program for whatever reasons (such as their teachers' interest in career education), then the population which those students represent is identified by describing the group's characteristics (e.g., low-income rural kindergarteners). A valid way of evaluating such a program would be to measure its effects on the students who were "selected" to participate in it and to describe precisely what the program consisted of and which national population the students represent. Then the program's potential adopter can predict its effectiveness for members of his/her student body who represent its same population. Of course, the larger the population to whom the results can be generalized, the more useful are the results to the more school systems. That is, a program which is found effective for a group of students representing all achievement levels may be more appealing than a program which has demonstrated its effectiveness only for students of below average achievement.

It is widely believed that in order for career education results to be generalized, the students involved in testing must be selected randomly from the school system implementing career education. This is true if the *population* of interest is students in the system. If the population is *national,* students can be chosen randomly from within the system but the sample is still far from a random sample of the national population; for all practical purposes, this sample would be no more random from a national perspective than a very deliberately chosen group within the school system. We do not wish to imply that every conceivable sampling technique will yield valid evaluation results, but rather that the practice of testing randomly-selected students from the school district or from schools or classrooms "participating" in the career education program is not the *only* valid sampling plan for career education evaluation.

This sampling strategy has, nonetheless, gained a great deal of popularity in career education evaluation, perhaps because of career education's intent to serve all students and because career education is more often infused into the curriculum than taught as a separate course. Let us examine what evaluation results tell us if we test, say, 50 sixth-graders selected at random from four elementary schools participating in the district's career education program. We will assume that a school's participation means that all of its teachers have received some minimum amount of inservice training and that the project's material and human resources are available to the school. We still cannot assume that the 50 children have experienced comparable career education. For that matter, we should not assume that all 50 children have experienced career education of any amount or type; to do so would be to ignore the widely-observed tendency for some teachers to reject the concepts of career education while others embrace it enthusiastically, for some teachers to emphasize self-awareness while others concentrate on career information, for some teachers to rely heavily on class discussions while others prefer field trips. Since the only reasonable assumption we can make about the career education experiences of these 50 students is that they probably differed a great deal from student to student, the results of the evaluation would tell us very little about the effectiveness of any particular set of *student* activities. What the results *would* indicate is the changes which can be expected in students whose teachers have

access to the project's resources. That is, it is not really a student program which is being assessed. Rather, the evaluation results are an indirect measure of the educational reform resulting from staff development and other project strategies.

This sampling approach has its value under certain circumstances. The questions these evaluation results answer are, "What has the career education project done for our school system's students?" and, "If another district adopts our staff development and supportive services systems, what student benefits can be anticipated?" The reason this sampling approach is listed here as a common weakness of career education evaluations is that it is often applied prematurely, before a more basic question is answered: "Does our newly-developed student program work?", or, "Are these really the experiences we should be asking educators and community members to provide for students?"

We cannot determine whether the program works unless we know that it has been taught to the students who are selected for testing the program's effectiveness. To make sure that tested students have experienced the program we could begin by selecting students to participate in the piloting of a new program and then measure its impact on them. Another approach that is often more practical, especially for projects where the individual teacher (or counselor, student, etc.) is left a good deal of latitude in designing the student program, is to select for testing a group of students who have experienced the program which most clearly resembles the project's model. If the model program is a separate course, sample selection is a simple matter of identifying the course's enrollees and testing either all or a random sample of them. If the model program is a form of classroom infusion, the evaluation sample usually consists of the students of teachers who have implemented the infusion curriculum to the greatest degree. Of course, students selected in this manner still will not have experienced exactly the same set of career education activities, even if they are drawn from a single classroom, but variations in their experiences will be much smaller than those of students drawn randomly from "participating" schools. The results of evaluations involving only "identified as treated" students, like those of evaluations using random selection, are generalizable to the population which the sample represents and tell us what impact the model program has *where it is applied* rather than the impact of the availability of resources for

applying the program, where those resources may or may not be utilized. The distinction between the two sampling approaches looses its meaning if all students in participating schools receive comparable career education instruction, but this appears to be a relatively rare circumstance.

In *What Does Career Education Do For Kids?*, 569 student outcome studies are synthesized, where a "study" is an assessment of one outcome objective at one grade level. Only 169 of these studies employed "identified as treated" sampling plans, even though most of the projects generating these studies were in relatively early stages of program development, where the priority question should concern the viability of their student programmatic models. It should be no surprise that these studies were three time more likely than studies employing random selection to show statistically significant positive results (60 + 169 = 36% vs. 48 + 400 = 12%). These results demonstrate one advantage of evaluating programs according to their impact on students who are known to have experienced them.

*Research Designs**

Over-reliance on national norms. One-time testing of a single group of students who are compared to national norms may result in useful needs assessment data, but the results are of very limited utility in evaluating the impact of a program. Drawing conclusions from results such as "after career education, the students scored significantly higher than the national mean" requires assuming that without career education, the evaluation sample's mean would be the same as the national population's. This is a questionable assumption for two reasons.

The first is that "national norms" are not really national. Some standardized tests are normed on larger and more representative samples of national populations than are others, but all are subject to errors in estimating the population's score distributions.

---

*Many of the points made here and in a later section on statistical analysis are addressed more thoroughly in the excellent publication: *A Practical Guide to Measuring Project Impact on Student Achievement*, by D. Horst, K. Tallmadge, and C. Wood of RMC Research Corporation, 1975, ERIC number ED 106376.

The second problem is that there is no reason to expect any given sample of students to match a national population unless the sample is drawn randomly from the national population. Even if every third-grader in a school district is tested, their mean score is likely to be significantly different from the fiftieth percentile because neither the community nor the local educational system is likely to be "typically American" in every relevant way. If the evaluation sample is fairly small or if it is not representative of the district, the chances of its matching the national population are reduced further.

However, norm-referenced tests have advantages when used in other evaluation designs where there is no need to assume that the students' scores would match the national norm without career education, such as in studies involving pre-post without comparison group data. This use of norms still involves assumptions concerning non-career-educated students' expected post-test performance relative to their pre-test performance but these relative assumptions are generally more defensible than absolute ones. This is somewhat parallel to the generalizability of evaluation results, in that the educator in California can reasonably expect the program which resulted in improved student learning in Arkansas to result in improvements in his/her school system even though the actual test scores of the California and Arkansas students are likely to differ to a statistically significant degree.*

Excessive use of the pre-post without comparison group design. If the objective of the evaluation is to achieve statistical significance, this is the design to use. But if the objective is to determine the impact of a career education program, other possibilities should be considered.

Reasonable cause-and-effect inferences can be drawn from studies of

---

*Incidentally, if a local sample is compared to a national norm, the Z, not the t, should be used to test whether the sample mean is significantly different from the population's. Since the t for independent groups compares two samples rather than a sample to a population, its use implies a rejection of the hypothesis that the norm represents the population's score distribution. Although it may be valid to reject that assumption, this rejection also says that the validity of the evaluation strategy is rejected by the evaluator.

pre-post growth under any one of the following circumstances:

1) Other research indicates that growth is unlikely without special intervention. For example, results presented in *What Does Career Education Do For Kids?* suggest that self-esteem is not likely to improve over a period of six or so months even with career education, so a pre-post without comparison design would be a conservative one for evaluating this outcome.

2) Logic indicates that growth is unlikely without intervention. For example, job-seeking skills can be considered to be of two types. Some of these skills, such as the preparation of a resumé, are purely cognitive and it is reasonable to assume that they will not be learned without special efforts. Interview skills, on the other hand, consist largely of poise and communication skills which could be acquired through routine experience.

3) The time between pre- and post-testing is extremely short. For example, the design would probably suffice for a two-week mini-course in career education.

We recognize the difficulty of securing the cooperation of the comparison groups needed for the more conclusive research designs. Nonetheless, the problem most often cited as a rationale for designs without comparison groups is that all students in the district have been exposed to career education either directly or indirectly. This is a problem, certainly, but its magnitude often seems to be exaggerated.

It is true that if some students in a school system experience career education, all are likely to receive some amount via their friends, bulletin boards, small changes in their teachers' practices resulting from contact with involved teachers, etc. But to say that this minimal career education exposure destroys these students' utility as a comparison group is to say that the minimal experience is equally or nearly as effective as an intensive, high-quality career education program. Accepting that assumption puts us in a poor position to ask educators and the community to devote significant efforts to the implementation of career education programs.

Only if career education is well-established in the district and virtually every student or every student in an identified sub-population (such as gifted children) is involved to a large degree should it be impossible to identify a reasonable comparison group within the district.

It may be necessary to settle for comparing the effects of a lot of career education to the effects of a little rather than to the effects of none, but if the more intense program is more effective and the evaluation is designed to be sensitive to the difference, this is not a serious compromise. It is much easier to identify a comparison group if the "identified as treated" sampling approach is used than where students are selected at random from participating schools. Often, students of teachers who have participated in some staff development activities but who have not become active in career education constitute adequate comparison groups.

Use of comparison groups whose credibility has not been established. The credibility of many evaluations could probably be enhanced through demonstrations of the comparison group's comparability to the career education group. The *Checklist* discusses several ways of doing this.

Descriptions of Evaluated Programs

Not specific. Although most final reports indicated in general terms the nature of the student programs whose impact were assessed, relatively few were specific about the kinds of activities comprising the programs. This seriously compromises the utility of evaluation results outside of and perhaps even within the school system.

Not quantified. Few evaluation reports provided insight into the question, "How much career education did it take to produce these student outcome results?"

Not matched to outcome results. Sometimes both treatment and outcome data were presented, but it was still impossible to identify the career education activities of each group of students whose outcomes were measured, and thus to determine which set of activities was associated with each set of outcomes.

It may often be artificial to identify for a single group of students the particular activities which produced each outcome. For example, a seventh-grade program may consist of field trips designed to convey career information, role-playing to clarify work values, and filmstrips to develop decision-making skills. However, explaining the evaluation results of each of these three outcomes only in the terms of the activity type which was designed with that outcome in mind probably would be overly simplistic and involve overlooking interactions among the three kinds of activities. That is, if success were demonstrated in decision-making skills but not in career information, one might conclude that the filmstrips

were effective but the field trips were not. But it could be that eliminating the field trips would reduce the program's impact on decision-making skills. The point is that whereas it is appropriate to discuss the rationale for each component of the student program, outcome results generally should be interpreted for the program as a whole for any given group of students.

However, the outcomes of seventh graders should be associated with the activities of the same seventh graders, particularly if the "identified as treated" sampling plan is used. Data indicating the average amount and type of career education received by all seventh graders in the school system will not suffice in interpreting the outcome results of students who were chosen for the evaluation because they had undergone a presumably above-average program. Similarly, treatment data which are aggregated across grade levels cannot be used to interpret the outcomes at any one grade level, regardless of the sampling plan. For example, knowing that K-12 students receive an average of three hours of career education instruction per week may be useful information, but it gives no insight into the amount of career education required to produce the accelerated reading achievement observed at second grade unless we assume that second graders received the same amount of career education as high school juniors and kindergarteners. As another example, knowing that 93 guest speakers visited elementary schools is in itself meaningless as treatment data. If a random selection sampling plan is used to evaluate fifth grade impact, the average number of visits to fifth-grade classrooms should be indicated. If three fifth-grade classrooms are chosen for evaluation in an "identified as treated" sampling plan, the treatment would best be described in terms of the number of guest speakers visiting each of the three classes.

Outcome Measurement Instruments and Strategies

The choice of measurement instruments is perhaps the single most important determinant of whether evaluation results will reflect the program's impact on students. Nonetheless, this decision often seems to receive less attention than it deserves. Career education outcome measurement poses considerable difficulties, but new instruments and strategies are emerging and the problem is becoming less serious than it was several years ago. Also, the option of locally-developed instruments should never be ruled out as a possibility; many programs have done this with great success.

In choosing instruments, the first question should be, "Does it measure what we want to evaluate?" This question must be answered locally, and by career education practitioners, not by evaluators alone. As a general rule, the best test is one where students who give "correct" answers represent what the program is attempting to produce. There are other considerations, of course, and expert opinion and psychometric data should be considered, but instruments chosen solely on the basis of how many other programs have used them or what "experts" far removed from the local program say about them are likely to result in disappointing evaluation results and invalid conclusions about the program's effectiveness.

A number of programs were made keenly aware of the need for careful selection of instruments after negative evaluation results were found and their examination of the tests revealed that they had little to do with the program's objectives. Some of them recommended, as do we, that instruments be scrutinized before, not after, their administration.

The rationale for measurement strategies should always be indicated in some way, but this is especially important if the technique is unusual. A few examples of cases where the rationale was not clear are:

* Even though career educators have emphasized time and again that they do not expect or desire young children to make career choices, three projects evaluated the wisdom of fourth-graders' career plans. In one case, even the rationale of the measurement technique was unclear--the student stated his or her ideal and realistic career choices and judges rated the difference between the two on a socio-economic scale, hoping that ideal and realistic career choices would show similar socio-economic status. We considered the measurement technique of the other two cases sound--judges rated the compatability of the student's top three career choices with his/her responses to each of thirty questions relating to personal interests, abilities and values--but assessing the validity of fourth-graders' career choices in any form still seems of questionable value.
* A program which emphasized the elimination of sex-role stereotyping employed an assessment technique which appears incompatible with the intent of such efforts. The test consisted of a list of

job titles, some male-stereotyped and some female-stereotyped. Students were asked to indicate whether each job was predominantly male, female, or dominated by neither sex. Results were "positive" in that students tended more toward "dominated by neither sex" after the program, but these results could be interpreted to mean that misinformation had been conveyed rather than that students had become more aware of and less subject to the inequities of the status quo.

* One project administered occupational interest inventories to middle school students on a pre-post basis and found that students' interests changed. It seems that this result could be anticipated with or without career education and the implicit assumption that interest changes are inherently desirable is debatable.

Although the direct approach to measurement is often the best, it can be overdone. Several programs asked students outright whether they had attained the program's outcome objectives; this was even done at third grade. Only slightly more subtle than asking students, "Do you know about a lot of different careers?" is the approach of asking students to report how much they know about each of a number of jobs.

Tests should measure the career education concepts which the program is designed to convey, but when the content of test items matches precisely the cotent of instruction, test results can become trivial. For example, one locally-developed test of occupational knowledge contained a photograph of a local business establishment which the career education students had visited. The fact that more career education than comparison students could identify the type of business that takes place in the building is not very convincing evidence that the field trip was worthwhile.

Inaccurate descriptions of what tests measure are very common, but the most popular is confusion between self-esteem and self-awareness. (e.g., "Self-awareness was measured with the Coopersmith Self-Esteem Inventory.") Self-awareness is an important career education outcome and probably the most challenging to measure, so we do not quarrel with the practice of measuring self-esteem in lieu of self-understanding. However, the limitations of measurement should be recognized, as overlooking them leads to faulty conclusions.

Some evaluations employed a single commercial test scale for evaluating impact with respect to two or more distinctively different objectives, such as decision-making skills and career knowledge. If the scale really measures both, it probably should not be used. If it measures one to a great degree and the other to a small degree, results should be considered an evaluation only of the dominant skill area, although the scale's description in the evaluation report should be accurate.

Some locally-developed tests addressed all of the program's outcome objectives through a single scale yielding a single score, making evaluation results difficult to interpret. Locally-developed tests usually have the advantage of being closely related to local objectives, but items addressing different objectives should be on different scales for evaluation results to indicate which objectives were achieved and which were not.

Measurement techniques for each of the USOE learner outcomes for career education are discussed in *What Does Career Education Do For Kids?* and all of the instruments used in synthesized student impact studies are listed in the appendix along with the number of studies in which each was used and the number where positive results were obtained. These data can be used as one consideration in choosing tests, as they provide a gross index of sensitivity to career education instruction. It should be remembered, though, that evaluation results are influenced by many factors other than the quality of the measurement instrument and that the results other programs demonstrate with a particular instrument should be only one of many considerations in instrument selection.

Statistical Analyses and Interpretation

No inferential analyses applied. Very often data collected from educators, community members, parents, etc., could have been analyzed inferentially in meaningful ways but were not. However, we do not see this as a very serious problem in most cases--the considerable effort involved in performing every reasonable analysis of all available data would be appreciated by few. On the other hand, it seems rather wasteful to launch massive student testing programs and then to report only mean scores of students before and after the program or of career education and comparison groups.

Vague references to statistical analyses. This weakness may fall into the category of incomplete reporting, but the problem could be more basic than that. Statements such as "differences were (or were not) significant," with no other discussion of inferential statistical analyses, raise doubts as to whether analyses were conducted. Statements such as "results were computer-analyzed," when made without further clarification, can leave the impression that an analysis was conducted but that the report writer does not know which one.

Poor choice of analyses. Almost any set of data can be legitimately analyzed in more than one way and often the analysis performed was not the one of greatest interest to the most people. In one case student test results were factor-analyzed but no t-test was applied to determine the significance of pre-post differences. Much more often, results were analyzed separately for each of several classroom units at a given grade level or differences between classroom units were analyzed, even though there was no indication that either the students or their career education experiences differed from class to class. If there is local interest in such analyses, there is nothing wrong with this practice, but a summary analysis including all of the career education students together should also be conducted for summary reports and rationale should be given for separate or comparative analyses by class.

Other poor choices of analyses were made in some cases where pre- and post-scores were available for both career education and comparison groups. The best use of these data is usually the analysis of covariance or a t-test on post-test results after a t-test on pre-test scores reveals no initial differences between groups. However, these tests were sometimes omitted even where several other t-tests were applied, such as matched-pairs t's for each group.

Misuse and misinterpretation of the analysis of covariance. The analysis of covariance is a useful statistic and very often the best one to use with pre- and post-data from career education and comparison groups. Nonetheless, it has been given more credit than is due.

The first and most serious fallacy is that it compensates for significant pre-treatment differences between groups. Unfortunately, it only capitalizes on the correlation between students' pre- and post-performance, thereby reducing the error term of F. This makes the analysis of covariance a very powerful statistic, but it does not eliminate the differences in expected growth rates of students who are drawn

from different pre-treatment populations and does not eliminate the necessity of choosing a comparison group which is comparable to the career education group. If the two groups differ significantly on the pre-test, a more elaborate regression technique is needed.

Another fallacy about the analysis of covariance is that it tests between-group differences in growth rates between testing dates, or that significant differences indicate that the career education group gained more than the comparison group. In reality, it tests differences between groups on the post-test, as does the t for independent groups; the difference is that post-test scores are adjusted for random pre-test differences and the analysis of covariance is more likely than the t to reveal significant differences. The misconception that the analysis of covariance tests for differences between groups' gains would not be so serious if it were not for the common belief that it also corrects for large initial differences between groups--if the pre-test means are equal, the difference between mean gains is equivalent to the difference between post-test means. However, the two misconceptions usually go hand-in-hand as in the case below where a t-test was applied to pre-test scores and its significance led to the decision to apply the analysis of covariance rather than the t to final results:

> The significant differences between the experimental and control groups exist in the areas of mathematics achievement and career knowledge. Since the final evaluation will be based upon gain scores, these differences are inconsequential.

Use of the wrong t-test. Just how often the t for independent groups was applied to pre- and post-scores of career education students is difficult to say because few reports specify which t was used. However, in cases where the N's associated with pre- and post-means were reported as unequal, it is reasonable to assume that the wrong t was applied.

The appropriate t for determining whether the difference between pre- and post-means of the same students is significantly different from zero is the t for matched pairs or correlated samples, which involves matching each student's pre-test score with his/her post-test score. Students who do not have scores on both tests should be excluded from this analysis. Using the t for independent groups on these data is invalid and is also less likely to reveal statistically significant results.

Confusion between statistical significance and casual relationships. Inferential statistics are nothing more than tools for determining whether observed differences are more likely to be real or random. They do not, in themselves, tell us whether career education is the cause of differences. Some evaluation designs allow more confidence that career education was the cause than do others, but we can never be certain that career education was the only, or even the major, factor influencing the results.

The most frequently over-interpreted statistical results are those of pre-post evaluation designs without comparison groups. For example, a discussion of the results of a matched-pairs t-test yielding a probability level of .001 went as follows:

> The question, of course, remains as to whether a comparable group without exposure to a career education program could have experienced like increases. Every study of which this writer is aware after ten years in the business of statistical inference suggests an answer of "no". Had the results of the study yielded a probability that such gains could have occured twenty, ten, even five times in one hundred by chance, this evaluator would be inclined to greater consciousness. But when the possibilities for chance to operate climb into the thousands against being operative, a rather firm conclusion evolves that the results were caused, and caused by the program's impact.

These evaluation results give us a high degree of confidence that the students' superior post-test performance was real, rather than the result of random score fluctuations. However, the probability level tells us nothing about the cause(s) of the improvement and it gives us no reason to believe that similar gains would not occur without career education. The magnitude of the gain, which is only one factor influencing the statistical significance level, would have been a more appropriate focus of the above discussion. That is, other research may have indicated that a raw score mean gain of 10 percent over the period of six months was remarkably high for this particular test, lending credence to the hypothesis that the improvement was not a simple function of the students' being six months older.

Confusion between statistical significance and educational significance. Statistical significance can be and often is achieved with a difference between means of one raw score point or less. Thus, the magnitude of the gain or of the difference between groups should be considered in terms of its practical significance to educational priorities.

The responsibility of deciding whether a given level of student impact justifies the maintenance, expansion, or replication of a program lies more appropriately with school administrators than with evaluators. However, evaluators can present the results in ways which facilitate these decisions; several suggestions on this topic are offered in the *Checklist*. Also, evaluators should avoid expressing excessive zeal over small but statistically significant differences, as this practice misleads some readers and leads other readers to question the evaluator's judgement.

PART II

UNIQUE SOLUTIONS TO COMMON EVALUATION PROBLEMS

Our review uncovered a number of high-quality evaluation strategies, many of them quite unique. Those which are most likely to apply to other programs are presented here as a demonstration of the growing capacity of career education evaluation and in hopes that they and equally creative evaluation strategies will be applied more frequently in the future. Student outcome measurement techniques are not included here, but are discussed in *What Does Career Education Do For Kids?*

Problem: As a result of their association with the project, the third-party evaluators have made observations and formulated recommendations which are not based on "hard" data. They want to share their observations in the final report, but the evaluation plan does not indicate this as the evaluators' role.

Solution: The evaluators prepare two final reports, one based on objective data and another clearly identified as informal, subjective observations of the project's functioning and results. (9)*

Problem: A simple system is needed for showing the project's activities in chronological order and for relating them to the operational plan of the funding proposal.

Solution: A chart identifying each major activity and its planned and actual start and completion dates. (2)

Problem: Counselors, teachers, and administrators are all trained together. Is the staff development program weighted in favor of one of these audiences, or is it equally well-received by all groups?

Solution: Participants rate the quality of the program on a five-point scale for each of several questions. The mean rating of each group is computed for each question. The one-way analysis of variance is applied to determine whether the groups' opinions of the program differed. The same type of analysis was applied to questions concerning knowledge of and attitudes toward career education. (3)

---

*Refers to list of referenced projects following this section.

Problem: A sample of teachers in each of several schools is asked multiple-choice questions concerning their involvement in career education and their perceptions of the project's impact on the school. Are there significant differences among schools?

Solution: Chi-square analysis of the number of individuals in each group giving each response. (11)

Problem: Measuring incremental quality improvement of a career education program in terms of its administrative structure and the support of various groups.

Solution: An incremental quality improvement model specifying five stages of "organizational infusion," a process through which externally-funded projects are institutionalized to the point where the district or school assumes total responsibility for the program's continuation and further development. The model is based on previous research concerning the factors influencing the long-term success of innovative programs and specifies standards for each stage of improvement. (1)

Problem: Identifying changes in the amount of career education being taught now with respect to the amount taught before the project began when no baseline data were collected.

Solution: Ask teachers (counselors, etc.), how their current involvement compares to their previous involvement in career education. (8) In this case teachers were asked to compare their present involvement with students to that of three years ago for each program component. For example:

> *Self Awareness:* Awareness of self (and others) as individuals who have certain likes and dislikes, abilities and disabilities, feelings, and values.
> a. Very much higher than 3 years ago
> b. Higher than 3 years ago
> c. About the same
> d. Less than 3 years ago

Collecting implementation data before and after the program is preferable, but this approach is reasonable in cases where baseline data have not been or cannot be obtained.

Problem: A pre-post with comparison group design is desired for student impact assessment, but it is impossible to predict at the beginning of the year which students will and will not become involved in the program.

Solution: Administer the pre-test to a large number of students. Toward the end of the year identify the career education group and the comparison group and administer the post-test. If the groups consist of the most and least-exposed students, some who were pre-tested need not be post-tested. (4)

Problem: At the secondary level some teachers and counselors are highly active in career education while others are not. Therefore, most students receive some career education but the amount varies a great deal from student to student. How can the amount of career education be measured and how can its effectiveness be evaluated?

Solution: Administer a student questionnaire yielding an interval or ratio-level measure of the individual's amount of career education experience. Calculate the Pearsonian correlation of the treatment measure with the outcome measure and test the significance of r. If r is significantly greater than zero, a statistically significant relationship exists between the amount of career education a student experiences and his/her test scores. (11)

This design is appropriate where student exposure to career education is highly variable and normally distributed and therefore where classifying students into career education and comparison groups would be artificial. If several treatment measures would be meaningful (such as the number of shadowing experiences, group counseling sessions, etc.), multiple regression analysis can be used to test their combined effects and also to determine which activities are most highly correlated with the outcome measures.

Problem: Getting enough information out of student test results for use in improving career education instruction.

Solution: Analysis of response patterns to each test item (7). Compute the proportion of career education and comparison groups answering each item correctly. Establish a standard of acceptable performance, such as 75 percent of students answering correctly. Divide the test items into four groups:

1. Both groups performed well. These items represent concepts which apparently are acquired without career education and eliminating these concepts from the career education curriculum may be desirable.

2. Neither group performed well. These items represent concepts which are not being acquired through the program and which require further emphasis.

3. Comparison students performed well and career education students performed poorly. If any items fall into this category, attempts to convey these concepts may have been counter-productive and the curriculum is probably in need of considerable revision.

4. Career education students performed well and comparison students performed poorly. These items represent concepts which are not acquired without career education and which are being conveyed successfully through the present career education curriculum.

This type of analysis is valid only if there is very good reason to believe that the career education and comparison groups would perform equally well without the program. Analyzing test results in this manner is especially important for test scales measuring a large variety of career education concepts but may be useful even for tests addressing a single objective such as career information.

Problem: Establishing group equivalence and applying powerful statistical analyses when no pre-test scores are available.

Solution: Use of aptitude or achievement test scores or grade point average as the covariate in the analysis of covariance (10). Any measure which correlates highly with post-test scores but which is not affected by the career education program is appropriate for use in the analysis of covariance model. If there is a possibility that career education may affect the covariate measure (such as reading achievement), the measure should be taken prior to the program. Since appropriate covariate measures are often available from school records, this evaluation strategy is an excellent approach when the evaluation is begun after the program is in progress.

Problem: Is the curriculum effective regardless of who teaches it to whom?

Solution: A two-way analysis of variance design, with time (pre-post) as a within-subjects variable and class as a between-subjects variable (6). Each class experiences the same curriculum, but they may be different in terms of teacher and student characteristics. The results of each effect are interpreted as follows:

Time: Does student performance improve?

Class: Are the classes different?

Time x Class: Does the amount of improvement vary from class to class?

Problem: Demonstrating the cost-effectiveness of a placement program.

Solution: Analysis of the tax dollars generated from the employment of students placed through the program (5).

REFERENCES

Project Title, Director and Address	Funding Source & Grant Number
1. I-ECC Career Education Incremental Improvement Project K-12 Lucinda L. Kindred, Project Director Industry-Education Council of California (I-ECC) 1575 Old Bayshore Highway Burlingame, California 94010	OCE G007503735
2. I Believe in Kids Albert Thomas, Jr., Project Director Jefferson County School Board P.O. Box 499 Monticello, Florida 32344	OCE G007502294
3. Illinois Career Education Area Service Centers (Urban and Rural): A Vehicle for Demonstration Carol Reisinger, Project Director Illinois Office of Education 100 North 1st Street Springfield, Illinois 62777	OCE G007503404
4. Career Resource Project Joe W. Roth, Project Director Indiana State Board of Vocational-Technical Education 401 Illinois Building 17 West Market Street Indianapolis, Indiana 46204	Part D 0-73-5312
5. An Exemplary Program for Career Education John J. Vandersypen, Jr., Site Coordinator Natchitoches Parish School Board P.O. Box 16 Natchitoches, Louisiana 71457	Part D 0-73-5308
6. New York State Consortium for Career Education Dr. Gordon E. Van Hooft, Project Director New York State Education Department Albany, New York 12234	OCE G007502353
7. D.E.C.E.M.--District Eleven Career Education Model Mrs. Wimell Thomas, Project Director Community School District #11 1250 Arnow Avenue Bronx, New York 10469	OCE G007503732

Project Title, Director and Address	Funding Source & Grant Number
8. Comprehensive Career Education Process in Springfield Public Schools Donovan D. Kimball, Project Director Springfield School District 525 Mill Street Springfield, Oregon 97477	Part D 0-73-5288
9. Language Experience Based Awareness + Hands On Exploration + Competency Based Preparation = A School Based Total Career Education Model Edward H. Lareau, Jr. & Clifford A. Bayliss, Jr. Admiral Peary Area Vocational-Technical School Rt. 422 W., R. D. #2 Edensburg, Pennsylvania 15931	Part D 0-73-5272
10. Career Education for Gifted and Talented Students William W. Cox, Project Director Highline Public Schools #401 15675 Ambaum Blvd., S.W. Seattle, Washington 98166	OCE G007502316
11. Highline's Career Alternatives Model Dr. Ben A. Yormark, Project Director Highline School District #401 15675 Ambaum Blvd., S.W. Seattle, Washington 98166	Part D 0-73-5289

PART III

A NOTE TO PROJECT DIRECTORS ON HOW TO GET THE MOST FROM YOUR THIRD PARTY

A third-party evaluator or evaluation agency can be a real asset to a program, but they can also be an expensive nuisance. Which experience you have depends largely on the evaluator(s) you select and on the way you work with him/her/them.

A common method of selecting an evaluator is to issue a request for proposals asking bidders to write an evaluation plan based upon the project's funding proposal or an abstract of it. The successful bidder is the one who seems to propose a reasonable plan at a low price. This approach has its merits, but it also has serious disadvantages.

The best evaluation plan is one which addresses the information needs of various audiences of a project through realistic means. Funding proposals tend to be less than crystal clear about the information needs of the project staff and local decision-makers, even though bidders may be able to identify the evaluation questions of interest to potential adopters of your model program. Also, evaluation designs are virtually always compromises of the ideal (from a research point of view) with the feasible (taking into consideration local constraints on data collection). Thus, the evaluator needs answers to such questions as, "Can we get the cooperation of a comparison group?" and, "Can we count on teachers to record their career education activities?" *before* preparing an evaluation plan. Since your help is needed in planning the evaluation, this step should follow, not precede, the selection of an evaluator.

The selection of evaluators should be taken as seriously as the selection of key members of your staff and should be handled in much the same way. First decide what kinds of services you expect. Will you need an instrument developer? A statistician? A management consultant? A deft report writer? An interviewer? A cost analysis specialist? Do you expect your evaluator to have experience in a particular evaluation technique? To be well-versed in career education? Most evaluators are good in at least one of these, but few individuals (or even agencies) excell in them all. Next, design your request for proposals in a way that will show which bidder has the best capacity for delivering the services you need. Copies of evaluation reports and evaluation plans the bidder has prepared for other programs can be elucidating. Another possibility is to ask the bidder to suggest alternative solutions to a particularly thorny evaluation problem you foresee, such as how to

measure the amount of career education high school students are getting.

Ask bidders for complete lists of their recent clients and follow up on all of them. Your telephone bill will probably be money well spent.

Evaluation services are expensive and you should be prepared for that, but there are steps you can take to increase your chances of getting what you pay for. Ask for bidders' fee schedules, of course, but also investigate how charges are computed. Otherwise, you could end up paying for a day's services for two hours of work. Or you and another client may both be charged in full for developmental work that applies to both projects. Contractors have even been known to charge two projects for the full travel cost of a single trip.

Project directors usually know how much they can afford to spend on evaluation, so asking for bids *per se* is to little advantage; it is better to find out how much quality, as well as quantity, each bidder can deliver for the price you have in mind. A good rule of thumb for budgeting a thorough external evaluation is ten percent of the total grant award. If this cannot be arranged, plan to perform some evaluation tasks internally, perhaps using the external evaluator more as a consultant and auditor than as a developer, data collector, data analyzer, and reporter. Remember, too, that doubling an evaluation budget more than doubles the services the money can buy, as there are certain fixed costs such as travel, planning time, and report preparation which vary relatively little with the amount of the contract.

Once an evaluator is chosen, a written and notarized contract is a good idea. Settle for a grant arrangement only if you have the utmost confidence in the integrity and reliability of the individual or agency you have chosen. It is usually preferable to insist on an accounting of all charges and to pay for services only after they are delivered. Make sure that the contract covers the possibilities of the contractor's over- or under-expending the budget and stipulates that the final payment will be withheld until the final evaluation report is submitted in acceptable form. If the individual or agency has a reputation for tardiness you may want to consider a stipulation that the contractor pay the project for every day between the final report's deadline and your receipt of it.

The earlier the signing of the contract, the better. Delays in instituting evaluation systems seriously compromise the quality of the final evaluation and also leave the program without important feedback systems when they are most needed, during the early stages of development. It is best, in fact, to line up an evaluator when the funding proposal is written, particularly now that the USOE is often switching to October start-up dates.

Time should be allowed for a lengthy meeting with your evaluator(s) as early as possible for planning the evaluation. He/she/they by then should be familiar with your proposal, but will need more details about the project's plans and priorities and about circumstances in the schools and community which will affect the evaluation.

Ask for a written draft evaluation plan soon thereafter. If you are not totally satisfied with it, negotiate revisions for the final version. The plan should specify the objectives, questions, or hypotheses to be addressed, the evaluation strategies associated with each, a time schedule, and who is responsible for what. Written plans are good for assessing the design's adequacy before it is too late to change it, for preventing misunderstandings, for facilitating the preparation of final reports, and for pacifying federal project officers.

Program evaluation designs can only be as systematic as the program itself. If the program's objectives are nonexistent or nebulous do not expect the evaluator to know what outcomes to look for. If plans for achieving the objectives are ill-defined or if they change daily, do not expect the evaluator to determine which strategies are successful. If the project staff do not know which schools and individuals are involved in project activities, do not expect the evaluator to know who to ask for evaluation data.

Always insist on reviewing data collection instruments before they are put into use, whether they are commercial or developed by your evaluator. Certainly the evaluator's advice concerning the technical soundness of an instrument should be given careful consideration, but the project staff should make sure that the instrument addresses the program's objectives and/or information needs.

Although this applies to all kinds of data collection instruments, it is particularly key for student tests. For example, careful examination of a test of decision-making skills may reveal that the test developer had a different definition of decision-making skills than does your program. If students who give all "correct" answers is not what your program is trying to produce, the results of the test may lead to unjustified conclusions about your program's effectivenss, and you have every right to veto its use. At the same time, though, realize that many career education outcomes are difficult to measure and that you are not likely to find or to be able to develop the "perfect" test for your program.

Evaluators cannot do their jobs without a great deal of cooperation on the part of the project staff, particularly in the area of data collection. If the mutually-agreed-upon evaluation plan requires the maintenance of project records, agree upon a system that will minimize the burden but be sure that the system is applied conscientiously. If the project staff agrees to take the responsibility of distributing and collecting questionnaires, follow through on the agreement. Incomplete data not only gives evaluators headaches; it also compromises the quality of your evaluation and usually costs you money for the extra time it takes your evaluator to compensate for the problem.

Communicate regularly with your evaluator. Frequent site visits are desirable, but if they can't be arranged you should talk on the phone and/or correspond at least monthly. Ask for prompt feedback on data collected and observations made; there is no need to wait until the final report to learn of preliminary evaluation findings. Similarly, keep the evaluator informed about the program's progress, problems, changes in plans, etc. He/she/they will need to know these things because they may affect the evaluation or its results in ways you may not foresee. Besides, evaluators have been known to give good advice on occasion.

When final report time draws near, meet with the evalutor to plan coordinated performance and evaluation reports. Our review showed that the two reports tend to be redundant, yet much important information is left out of both. Both of these problems can be avoided by outlining the reports together.

If at all possible, ask for a draft final evaluation report well in advance of the deadline for submitting the final performance report. Be reasonable, of course, and do not expect it three days after the last data are collected, but barring vacations and unusual circumstances a month should be plenty of time. If you feel the report is incomplete or unfair, ask for revisions, keeping in mind that the evaluation budget is probably by now expended and that it is unfair to ask evaluators to violate their own integrity.

We consider project directors not only justified, but duty-bound to request corrections of errors such as: evaluators' opinions presented as facts, misinformation (such as incorrect reports of the number of schools the project serves), interpretations of data which do not take into account significant factors of which the evaluator was unaware, and reports or sections of reports which make no sense either because they are badly written or because they are lacking in important information.

Revisions which should *not* be requested are changes in objective evaluation results, omissions of data which are not complementary to the project, and eliminations of critical but substantiated comments about the project.

Evaluators often find themselves in the curious position of being obligated to bite the hand that feeds them. With an appreciation of this situation and other problems evaluators face balanced by a recognition of the rights of consumers of evaluation services, the project director can do much to make the evaluator a key contributor to the program's success.